



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

<p>(51) International Patent Classification ⁶ : H04B 10/105</p>	<p>A1</p>	<p>(11) International Publication Number: WO 99/21301</p> <p>(43) International Publication Date: 29 April 1999 (29.04.99)</p>
<p>(21) International Application Number: PCT/US98/21625</p> <p>(22) International Filing Date: 16 October 1998 (16.10.98)</p> <p>(30) Priority Data: 08/953,424 17 October 1997 (17.10.97) US</p> <p>(71) Applicant: LOCKHEED MARTIN MISSILES AND SPACE [US/US]; 1111 Lockheed Martin Way 26-02/101, Sunnyvale, CA 94089 (US).</p> <p>(72) Inventors: OGASAWARA, Gary, H.; 786 Gull Avenue, Foster City, CA 94404 (US). JU, Timothy; 1139 Amarillo Avenue #2, Palo Alto, CA 94303 (US).</p> <p>(74) Agent: PENNINGTON, Edward, A.; Morgan & Finnegan L.L.P., 345 Park Avenue, New York, NY 10154 (US).</p>		<p>(81) Designated States: European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).</p> <p>Published <i>With international search report. Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.</i></p>
<p>(54) Title: DECISION-THEORETIC SATELLITE COMMUNICATIONS SYSTEM</p> <div data-bbox="337 1171 1266 1654"> <pre> graph LR 20([Demand Info]) --> 22([Network Performance]) 21([Supply Info]) --> 22 23[Allocation Decisions] --> 22 22 --> 24{Utility} </pre> </div> <p>(57) Abstract</p> <p>A communications system allocates communication bandwidth to calls in such a manner for optimizing network system performance (22). The system uses knowledge about the state of the past and present network demand (20), and network resources. Predictions of future demand and future network resources are used to calculate future network performance (22). Allocation decision (23) are chosen based on network performance (22) calculations and probabilities of achieving future network states.</p>		

© 1994 by the International Labour Office. All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording, or by any information storage or retrieval system, without prior permission in writing from the International Labour Office.

ü ü ü ü 卩 .δ .√ °≥ .σ θ≡ ² ■ ≈ .√ .■ ∫ °° ²n ° ∫ .■ °≡ .√ °■ .² °√ Γ■ €≡ ≈²

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon			PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

DECISION-THEORETIC SATELLITE COMMUNICATIONS SYSTEM

BACKGROUND OF THE INVENTION

1. Field of the Invention

The present invention relates to the field of telecommunications. More particularly,
 5 the present invention relates to a system and a method for allocating communication system
 resources to calls for optimizing network system performance.

2. Description of Related Art

In a typical satellite communications system, caller communication units are linked
 10 to ground-based communication facilities, or ground nodes, which, in turn, are linked to
 satellites. In some cases, individual communication units are linked directly to a satellite.
 The amount of caller traffic in a communications system is limited by the available
 communication bandwidth of the system. When caller demand exceeds the capacity of the
 system, new callers are denied access.

15 In conventional communication systems, there is a fixed amount of satellite
 bandwidth assigned to each ground node. The available bandwidth in these systems is
 often used inefficiently. For example, one ground node may be overloaded with calls,
 denying access to new callers, whereas a neighboring ground node may be underutilized,
 leaving unused satellite bandwidth assigned to that ground node. When a ground station
 20 becomes overloaded with calls, there is no mechanism in a conventional satellite system for

transferring satellite bandwidth from an underutilized ground station.

To overcome this deficiency, demand-assigned multiple access (DAMA) systems utilize a central control station that allocates satellite bandwidth on a call-by-call basis.

Such conventional systems employ a simple, greedy strategy for assigning satellite
5 bandwidth. That is, a ground node requests additional bandwidth from the central control station when a new call is received. The central control station assigns satellite bandwidth to the ground nodes as bandwidth requests are received allowing bandwidth to be reassigned from an underutilized ground node to an overloaded one.

Instead of using a greedy approach, a few conventional satellite systems use a
10 predictive approach, whereby the central control station predicts which stations will be receiving a high call demand, and allocates extra bandwidth to those stations in anticipation of their future need.

Nevertheless, conventional greedy DAMA systems and predictive DAMA systems do not always allocate resources in an optimal manner. For example, in a system in which
15 calls have different priority levels, if available bandwidth is allocated for handling a high demand of low priority calls, future higher priority calls may be blocked, or calls in progress may require preempting for satisfying demands of higher priority calls. As another example, if the system allocates all available bandwidth to one ground node for handling a high call demand or a high predicted call demand, there might be a lack of
20 bandwidth available for another ground node that will receive high call demand in the

future.

What is needed is a system that measures trade-offs involved in allocating bandwidth between stations over time. Further, what is needed is a system that determines an optimal allocation of communication resources (i.e., bandwidth) that maximizes overall
5 network system performance.

SUMMARY OF THE INVENTION

The present invention provides a communications system that measures trade-offs involved in allocating bandwidth between stations over time. The present invention further
10 provides a communications system that determines an optimal allocation of communication resources (i.e., bandwidth) maximizing overall network system performance.

The advantages of the present invention are provided by a system that includes ground nodes receiving calls and a control unit allocating communication resources to received calls by maximizing a communication system expected utility function. The
15 control unit allocates communication system resources, such as communication bandwidth, by assigning unused communication resources to calls, blocking calls, or preempting calls in progress.

The control unit maximizes the expected utility function of the communication system by identifying decisions relating to communication resource allocation, each
20 decision d_i being a different communication resource allocation decision. The control

unit determines a maximum expected utility E as:

$$E[u(d_i, C)] = \sum_{s_j \in S} u(s_j, d_i) p(s_j | d_i, C),$$

wherein s_j is a possible configuration of communication bandwidth assigned to individual calls, C is a system context based on the past and/or present network state,

- 5 $p(s_j | d_i, C)$ represents a probability of achieving state s_j given a decision d_i and the system context C, and $u(s_j, d_i)$ represents a utility function based on a weighted combination of network performance characteristics. The control unit selects the optimal system resource allocation decision d_i , by choosing the decision that maximizes the expected utility E of the communication system.

- 10 The communication resources that are allocated are time slots, frequency channels, or a combination of time slots and frequency channels. The control unit includes a memory that stores information representing past caller demand and network supply (i.e., network resource availability) information. The past caller demand information can include ground node caller demand, call holding times, call
- 15 source and destination ground nodes, data rate requirements, error rate requirements, and call priority levels. The network supply information can include network connectivity, call delay times, satellite link reliability, ground node reliability, battery capacity of individual satellites, and relative location of individual satellites.

The utility function $u(s_j, d_i)$ is a weighted combination of network performance

characteristics which can include grade of service (percentage of blocked calls), throughput, message delay, utilization, number of satellite hops, percentage of calls preempted, quality of service, and satisfaction of higher priority call requests. The utility function $u(s_j, d_i)$ can also be a summation of a past utility function, a present
5 utility function, and a future utility function. Calls can have an associated priority level and can be either voice or data.

The control unit can allocate an optimal communication resource configuration for each call request, or the control unit can allocate an optimal communication resource configuration after a predetermined number of call requests are received, or
10 periodically after a predetermined time period has elapsed. Each call individual call can be divided into multiple packets. The control unit can make a resource allocation decision for the entire call, or for each individual packet.

BRIEF DESCRIPTION OF THE DRAWINGS

15 The present invention is illustrated by way of example and not limitation in the accompanying figures in which like reference numerals indicate similar elements and in which:

Figure 1 illustrates a satellite-based communications system;

Figure 2 is a diagram showing influences affecting a decision problem for
20 allocating communication system resources; and

Figure 3 shows details of an influence diagram relating past, present, and future network states for allocating communication system resources according to the present invention.

Figure 4 is a flow chart of the basic steps used by the system of Figure 1 to
5 allocate communication resources to call requests;

DETAILED DESCRIPTION

The present invention provides a Decision-Theoretic DAMA (DT-DAMA) technique that intelligently allocates bandwidth requests, call blocks, and call preemptions
10 based on a current network state and predictions of a future network state to maximizes overall network system performance. The DT-DAMA approach uses Bayesian decision theory to intelligently allocate communication resources to call requests. The system decides to either assign bandwidth to calls, block calls, or preempt calls in progress to maximize overall network system performance. Bayesian decision theory combines
15 probabilistic reasoning (predicting future bandwidth supplies and caller demands) with utility reasoning (measuring network performance characteristics according to predetermined system preferences).

Compared to neural networks or rule-based systems, Bayesian decision theory provides both accurate modeling of complex systems and efficient run-time execution.
20 Simple optimization functions like those often used by neural networks are often unable to

explicitly model complex problems, and because of the lack of system initialization data, the neural networks usually require extensive tuning and training to converge to an acceptable solution. Rule-based systems which encode general directives like: "If a request has routine priority AND there is available bandwidth, THEN grant the request" are
5 inadequate for complex and uncertain systems, where it becomes difficult to create and maintain an acceptable set of rules.

Figure 1 shows an exemplary satellite communications system that utilizes the present invention. The exemplary system of Figure 3 includes a plurality of individual communication units 11 and 12, a plurality of ground stations 10, a plurality of satellites 13
10 and a control unit 14 with a memory 15. Individual communication units 11 are linked in a well-known manner to a ground station 10 which, in turn, is linked in a well-known manner to at least one satellite 13. Individual communication units 12 can link directly to the satellite 13. A control unit 14 makes decisions for allocating communication resources to calls, such as communication bandwidth, so that network system performance is optimized.
15 Control unit 14 also contains a memory 15 that stores past caller and network resource history information. Control unit 14 may be located in a separate ground facility, as shown in Figure 1, but may also be located inside one of the ground stations 10, in a satellite 13, or distributed in a well-known manner among multiple ground stations and satellites.

Figure 2 is a diagram showing the communication network resource
20 communication network resource allocation decision problem as an influence diagram. In

an influence diagram, ovals represent uncertain variables, rectangles represent decisions, and diamonds represent utility variables. Arrows pointing to ovals represent conditional probabilistic dependence. Arrows pointing to diamonds represent conditional utility dependence. The lack of an arrow implies conditional independence.

5 The details of network demand information 20, network supply information 21 (i.e., network resource availability), and network performance 22 variables are abstracted into the three respective uncertain variables. The utility 24 of the decision problem is directly dependent only on the network performance characteristics. Allocation decisions rectangle 23 represents decisions to be optimized, that is, allocations of satellite bandwidth
10 over a duration of the analysis period.

Figure 3 provides a more detailed model of the same decision problem for satellite communications resource allocation. In this influence diagram, the model has been temporally divided into past 3a, current 3b, and future 3c states. For simplicity, only two network performance variables are shown: grade-of-service (GOS) and throughput. The
15 expected utility of a network state is a function of the network performance variables with arrows pointing into the diamond utility nodes. In the example shown, the utility functions are maximized when throughput and GOS are maximized. The final utility node is also a weighted function of the sub-utilities at each time point in the past, present, and future.

Information that is available is entered as evidence by setting the values of the
20 nodes in the influence diagram. For example, the information for the past nodes 3a is

generally known, and their values can be input into the corresponding node. Then, through inference (i.e., application of probability and utility axioms), updated probability distributions on the other nodes are determined. Let $V = \{v_1, \dots, v_n\}$ represent the n uncertain variables in the future nodes (future demand, future supply). Let $pa(v_i)$ be defined as the set of parents of v_i , that is, the past and current nodes with arrows pointing into v_i . The probability of achieving a certain future network state is then calculated by:

$$P(v_1, \dots, v_n) = \prod_{1 \leq i \leq n} P(v_i | pa(v_i)) \quad (5)$$

This process is repeated for each individual decision that is evaluated. The decision leading to the highest expected utility is the optimal choice.

Figure 4 is a flow chart depicting how control unit 14 allocates bandwidth to calls so that system performance is optimized. As calls are placed by communication units 11 and 12, control unit 14 receives requests for bandwidth for the calls, at step 40. Control unit 14 can allocate communication bandwidth as each individual call is received, after a predetermined number of calls is received, or periodically after a predetermined time period as elapsed.

At step 41, control unit 14 determines present network demand and network supply (i.e., network resource availability) information. At step 42, the control unit 14 retrieves past network demand and supply history information from memory 15. Examples

of variables that represent caller demand information used by the present invention include:

(1) call frequency for each ground node, (2) call holding times, (3) source and destination nodes for each call, (4) call priority levels, and (5) data and error rate requirements. Examples of network supply variables that can be used by the present invention include: (1) network connectivity, (2) call delay times, (3) satellite link reliabilities, (4) ground link reliabilities, (5) ground node reliabilities, (6) satellite locations, and (7) satellite battery capacities.

After receiving call requests and past and present network information, at step 42, control unit 14 determines n possible decisions for allocating system bandwidth to the call requests received in step 40. Control unit 14 determines which allocation decisions are possible by evaluating present network supply and demand information. Each individual decision $d_i \in D$ may consist of assigning unused bandwidth to calls, blocking calls, or preempting existing calls. Each decision results in a different configuration of assigning time slots and/or frequency channels to call requests. Calls can be broken up into packets, and the control unit 14 can make a resource allocation decision for each individual packet or for the entire call. Assigning communication bandwidth on the packet level frequently produces better performance due to the smaller granularity of the unit processed.

In step 44, control unit 14 selects one allocation decision d_i for allocating bandwidth to calls (or call packets), and based on the selected decision, calculates

probability distributions for future network supply and demand variables. For example, the control unit 14 calculates probability distributions for future overall caller demand, high-priority call demand, link reliability, etc. The control unit 14 calculates probability distributions based on the particular allocation decision d_i and past and present network supply and demand information, as shown by steps 41 and 42.

Once the probability distributions for future network supply and demand variables are calculated, control unit 14 measures overall network system performance based on the selected allocation decision d_i . Network system performance is measured by calculating a utility function that is a weighted combination of various network performance characteristics. The network performance characteristics, are functions of the network supply and demand information. Utility represents system preferences for achieving different network performance criteria. Examples of network performance characteristics: (1) grade of service (percentage of calls that fail to connect on the first attempt), (2) throughput, (3) message delay, (4) network utilization, (5) number of satellite hops, (6) percentage of calls preempted, (7) quality of service, and (8) satisfaction of higher priority calls.

A utility function can be a summation of a past utility function, a present utility function, and a future utility function. The past utility function measures past network performance. Similarly, present utility function measures present network performance, while the future utility function measures future network performance.

One example of a utility function that can be used by the present invention is a linear weighted summation of network performance characteristics, such as:

$$\begin{aligned}
 U = & A1 \text{ (Grade of Service)} + A2 \text{ (Throughput)} + A3 \text{ (Message Delay)} \\
 & + A4 \text{ (Utilization)} + A5 \text{ (Quality of Service)} + A6 \text{ (Number of Satellite Hops)} \\
 & + A7 \text{ (Percentage of Calls Preempted)} + A8 \text{ (Quality of Service)}
 \end{aligned} \tag{1}$$

In the example utility function of Equation (1), if throughput is more highly valued than grade of service, then coefficient A_2 would be greater than coefficient A_1 by the degree necessary to obtain satisfactory throughput in accordance with the relative valuing.

The future utility function is more difficult to calculate than the past and present utility functions because the future network supply and demand information is not known. Future network performance is measured by the use of an expected utility function. The expected utility function combines the probability distributions of future network supply and demand variables calculated in step 44, and the network performance criteria. An expected utility function is calculated using:

$$E[u(d_i, C)] = \sum_{s_j \in S} u(s_j, d_i) p(s_j | d_i, C) . \tag{2}$$

The function $p(s_j | d_i, C)$ represents the probability of achieving a future network state s_j given allocation decision d_i selected in step 44 and a system context C .

probability distributions for future network supply and demand variables. For example, the control unit 14 calculates probability distributions for future overall caller demand, high-priority call demand, link reliability, etc. The control unit 14 calculates probability distributions based on the particular allocation decision d_i and past and present network supply and demand information, as shown by steps 41 and 42.

Once the probability distributions for future network supply and demand variables are calculated, control unit 14 measures overall network system performance based on the selected allocation decision d_i . Network system performance is measured by calculating a utility function that is a weighted combination of various network performance characteristics. The network performance characteristics, are functions of the network supply and demand information. Utility represents system preferences for achieving different network performance criteria. Examples of network performance characteristics: (1) grade of service (percentage of calls that fail to connect on the first attempt), (2) throughput, (3) message delay, (4) network utilization, (5) number of satellite hops, (6) percentage of calls preempted, (7) quality of service, and (8) satisfaction of higher priority calls.

A utility function can be a summation of a past utility function, a present utility function, and a future utility function. The past utility function measures past network performance. Similarly, present utility function measures present network performance, while the future utility function measures future network performance.

One example of a utility function that can be used by the present invention is a linear weighted summation of network performance characteristics, such as:

$$\begin{aligned}
 U = & A1 \text{ (Grade of Service)} + A2 \text{ (Throughput)} + A3 \text{ (Message Delay)} \\
 & + A4 \text{ (Utilization)} + A5 \text{ (Quality of Service)} + A6 \text{ (Number of Satellite Hops)} \\
 & + A7 \text{ (Percentage of Calls Preempted)} + A8 \text{ (Quality of Service)}
 \end{aligned} \tag{1}$$

In the example utility function of Equation (1), if throughput is more highly valued than grade of service, then coefficient A_2 would be greater than coefficient A_1 by the degree necessary to obtain satisfactory throughput in accordance with the relative valuing.

The future utility function is more difficult to calculate than the past and present utility functions because the future network supply and demand information is not known. Future network performance is measured by the use of an expected utility function. The expected utility function combines the probability distributions of future network supply and demand variables calculated in step 44, and the network performance criteria. An expected utility function is calculated using:

$$E[u(d_i, C)] = \sum_{s_j \in S} u(s_j, d_i) p(s_j | d_i, C) . \tag{2}$$

The function $p(s_j | d_i, C)$ represents the probability of achieving a future network state s_j given allocation decision d_i selected in step 44 and a system context C .

System context C includes the past and present network demand and supply information. Each individual outcome state $s_j \in S$ represents a possible allocation of communication bandwidth time slots or frequency channels to individual calls (or packets). The probability of achieving each network outcome state s_j is calculated
 5 based on the probability distributions of future network supply and demand information calculated in step 44. The function $u(s_j, d_i)$ represents the utility of state s_j after decision d_i has been made.

As a simple example of how the expected utility function operates, consider a system where the only network performance criteria of interest is throughput.
 10 Suppose control unit 14 determines three outcome states: high throughput, medium throughput, and low throughput. Based on making a particular allocation decision d_i and the predictions of future network supplies and demands, control unit 14 calculates p_{high} , p_{med} , and p_{low} , the probabilities of achieving a high, medium, and low throughput, respectively. Control unit 14 also calculates u_{high} , u_{med} , and u_{low} , the utility or "value"
 15 that the system attaches to achieving high, medium, and low throughputs, respectively. The expected utility of making decision d_i is then:

$$E(d_i) = p_{high} u_{high} + p_{med} u_{med} + p_{low} u_{low} \quad (3)$$

After the expected utility E has been calculated for the first decision d_i , then control unit 14 returns to repeat steps 44 and 45 for each of the n allocation decisions.

In step 47, the system then selects the decision d^* where:

$$d^* = \operatorname{argmax}_{d_i \in D} E[u(d_i, C)]. \quad (4)$$

Decision d^* represents the decision which produces the maximum expected utility E . This is the allocation decision that produces the optimal network performance. At step 48, control unit 14 then allocates bandwidth to the calls (or call packets), as determined by decision d^* . Lastly, at step 49, control unit 14 stores the present network supply and demand information in memory 15, to be retrieved in future cycles for calculating updated probability distributions.

The expected utility E is large for decisions that result in high probabilities of achieving high-performance network states. As a simple example of overall system operation: if there is a relatively low probability for ground station A (Figure 1) to have a high caller demand during the next hour, then the control unit 14 is not likely to assign additional bandwidth to ground station A, because a decision to assign additional bandwidth to ground station A results in a low expected utility E . However, if ground station A is predicted to have a small amount of high priority calls during the next hour, control unit 14 may decide to assign additional bandwidth to ground node A when the utility function places enough value on satisfying high priority calls. If the utility function places more value on overall system throughput, rather than satisfying priority calls, then control unit 14 will not likely assign additional

satellite bandwidth to ground station A. In summary, whether ground station A is allocated bandwidth depends both on the probability that A will receive a high caller demand, and the value that the system places on satisfying that demand.

THE CLAIMS

What is claimed is:

- 1 1. A communications system, comprising:
2 a ground node receiving calls; and
3 a control unit maximizing a communication system expected utility function
4 and allocating communication resources to received calls based on the communication
5 system expected utility function.

- 1 2. The system according to claim 1, wherein, preferably, the control unit
2 allocates communication system resources by performing at least one of:
3 assigning unused communication resources to calls, the
4 communication resources including communication bandwidth,
5 blocking calls, and
6 preempting calls in progress.

- 1 3. The system according to claim 2, wherein the control unit maximizes the
2 expected utility function of the communication system by:
3 identifying decisions relating to communication resource allocation, each
4 decision d_i being a different communication resource allocation decision;

5 determining a maximum expected utility E as:

6
$$E[u(d_i, C)] = \sum_{s_j \in S} u(s_j, d_i) p(s_j | d_i, C),$$

7 wherein s_j is a possible configuration of communication bandwidth
8 assigned to individual calls,

9 wherein C is a system context based on at least one of a present
10 network state and a past network state,

11 wherein $p(s_j | d_i, C)$ represents a probability of achieving state s_j given
12 a decision d_i and the system context C,

13 wherein $u(s_j, d_i)$ represents a utility function based on a weighted
14 combination of network performance characteristics; and

15 selecting the resource allocation decision d_i that maximizes the expected
16 utility E of the communication system.

1 4. The system according to claim 3, wherein the communication resources
2 that are allocated are one of time slots, frequency channels, and a combination of time
3 slots and frequency channels.

1 5. The system according to claim 4, wherein the control unit includes a
2 memory storing information representing past caller demand and network resource
3 information.

1 6. The system according to claim 5, wherein the past caller demand

2 information includes at least one of caller demand of individual ground stations, call
3 holding times, locations of source and destination ground nodes of individual calls, data
4 rate requirements, error rate requirements, and priority levels of individual calls.

1 7. The system according to claim 6, wherein the network resource
2 information includes at least one of network connectivity, call delay times, satellite link
3 reliability, ground node reliability, battery capacity of individual satellites, and relative
4 location of individual satellites.

1 8. The system according to claim 7, wherein the utility function $u(s_j, d_i)$ is a
2 weighted combination of at least one of grade of service, percentage of blocked calls,
3 throughput, message delay, utilization, number of satellite hops, percentage of calls
4 preempted, quality of service, and satisfaction of higher priority call requests.

1 9. The system according to claim 8, wherein the utility function $u(s_j, d_i)$ is
2 further a summation of at least one of: a past utility function, a present utility function, and
3 a future utility function.

1 10. The system according to claim 9, wherein each call has an associated
2 priority level.

1 11. The system according to claim 10, wherein each call is one of a voice call
2 and a data call.

1 12. The system according to claim 11, wherein the control unit allocates an
2 optimal communication resource configuration for each call request.

1 13. The system according to claim 11, wherein the control unit allocates an
2 optimal communication resource configuration after a predetermined number of call
3 requests are received, by determining the maximum expected utility E of resource
4 allocation decisions for the predetermined number of calls.

1 14. The system according to claim 11, wherein the control unit allocates an
2 optimal communication resource configuration periodically, with a predetermined time
3 period, by determining the maximum expected utility E of resource allocation decisions of
4 at least one call in that time period.

1 15. The system according to claim 11, wherein each individual call is divided
2 into multiple packets.

1 16. The system according to claim 12, wherein each packet is assigned a
2 different communication resource allocation.

1 17. A method for allocating communication resources, the method comprising
2 the steps of:

3 receiving call requests from a plurality of communication units;
 4 identifying possible communication resource configurations for assignment
 5 to each call request, each communication resource including a communication bandwidth;
 6 determining an optimal communication resource configuration for each call
 7 request by maximizing an expected utility function of the communication system.

1 18. The method according to claim 17, wherein the step identifying possible
 2 communication resource configurations to each call request, includes at least one of
 3 assigning unused communication bandwidth, blocking calls, and preempting calls in
 4 progress.

1 19. The method according to claim 18, wherein the step of determining the
 2 optimal communication resource configuration includes the step of maximizing the
 3 following expected utility equation:

$$4 \quad E[u(d_i, C)] = \sum_{s_j \in S} u(s_j, d_i) p(s_j | d_i, C)$$

5 wherein s_j is a possible configuration of communication bandwidth assigned
 6 to individual calls,

7 wherein C is a system context of at least one of a present network state and
 8 a past network state,

9 wherein $p(s_j | d_i, C)$ represents a probability of achieving state s_j given a
 10 decision d_i and the system context C ,

11 wherein $u(s_j, d_i)$ represents a utility function based on a weighted

12 combination of network performance characteristics; and

1 20. The method according to claim 19, wherein the communication bandwidth
2 includes at least one of time slots and frequency channels.

1 21. The method according to claim 20, further including a step of storing past
2 and present caller demand and network resource information.

1 22. The method according to claim 21, wherein the past and present caller
2 demand information includes at least one of: caller demand of individual ground stations,
3 call holding times, locations of source and destination ground nodes of individual calls,
4 data rate requirements, error rate requirements, and priority levels of individual calls.

1 23. The method according to claim 22, wherein the network resource
2 information includes at least one of: network connectivity, call delay time, satellite link
3 reliability, ground node reliability, battery capacity of individual satellites, and relative
4 location of individual satellites.

1 24. The method according to claim 23, wherein the utility function $u(s_j|d_i)$
2 includes at least one of: grade of service, percentage of blocked calls, throughput, message
3 delay, utilization, number of satellite hops, percentage of calls preempted, quality of
4 service, and satisfaction of higher priority call requests.

1 25. A method according to claim 24, wherein each incoming call has an
2 associated priority level.

1 26. The method according to claim 25, wherein each call is one of a voice call
2 and a data call.

1 27. The method according to claim 26, wherein the control unit allocates an
2 optimal communication resource configuration for each call request.

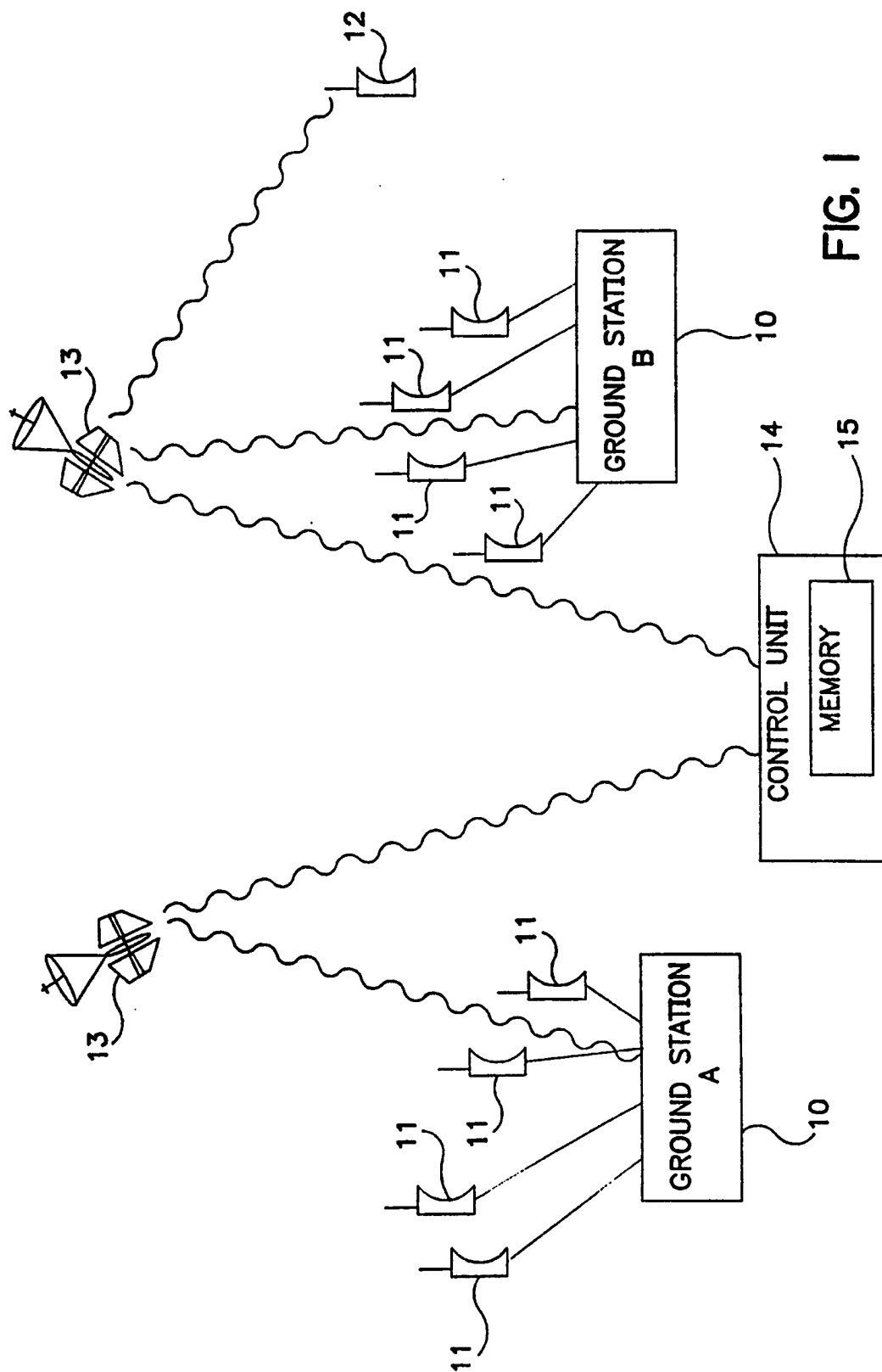
1 28. The method according to claim 26, wherein the control unit allocates an
2 optimal communication resource configuration after a predetermined number of call
3 requests are received, by determining the maximum expected utility E of multiple
4 sequential resource allocation decisions.

1 29. The method to claim 26, wherein the control unit allocates an optimal
2 communication resource configuration periodically, with a predetermined time period, by
3 determining the maximum expected utility E of resource allocation decisions of at least
4 one call in that time period.

1 30. The method according to claim 26, wherein each individual call is divided
2 into multiple packets.

1 31. The method according to claim 29, wherein each packet is assigned a

- 2 different communication resource allocation.



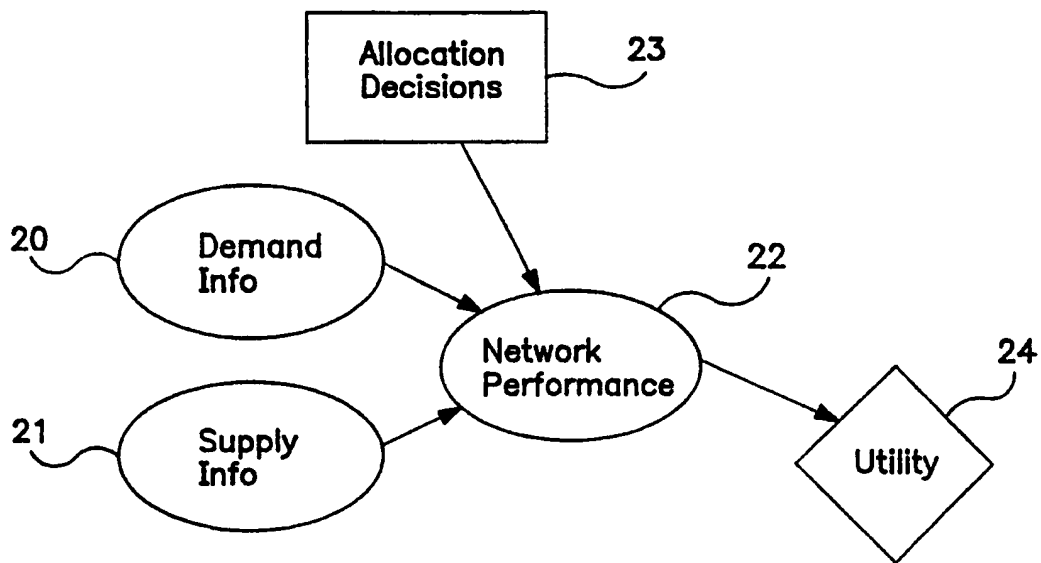


FIG. 2



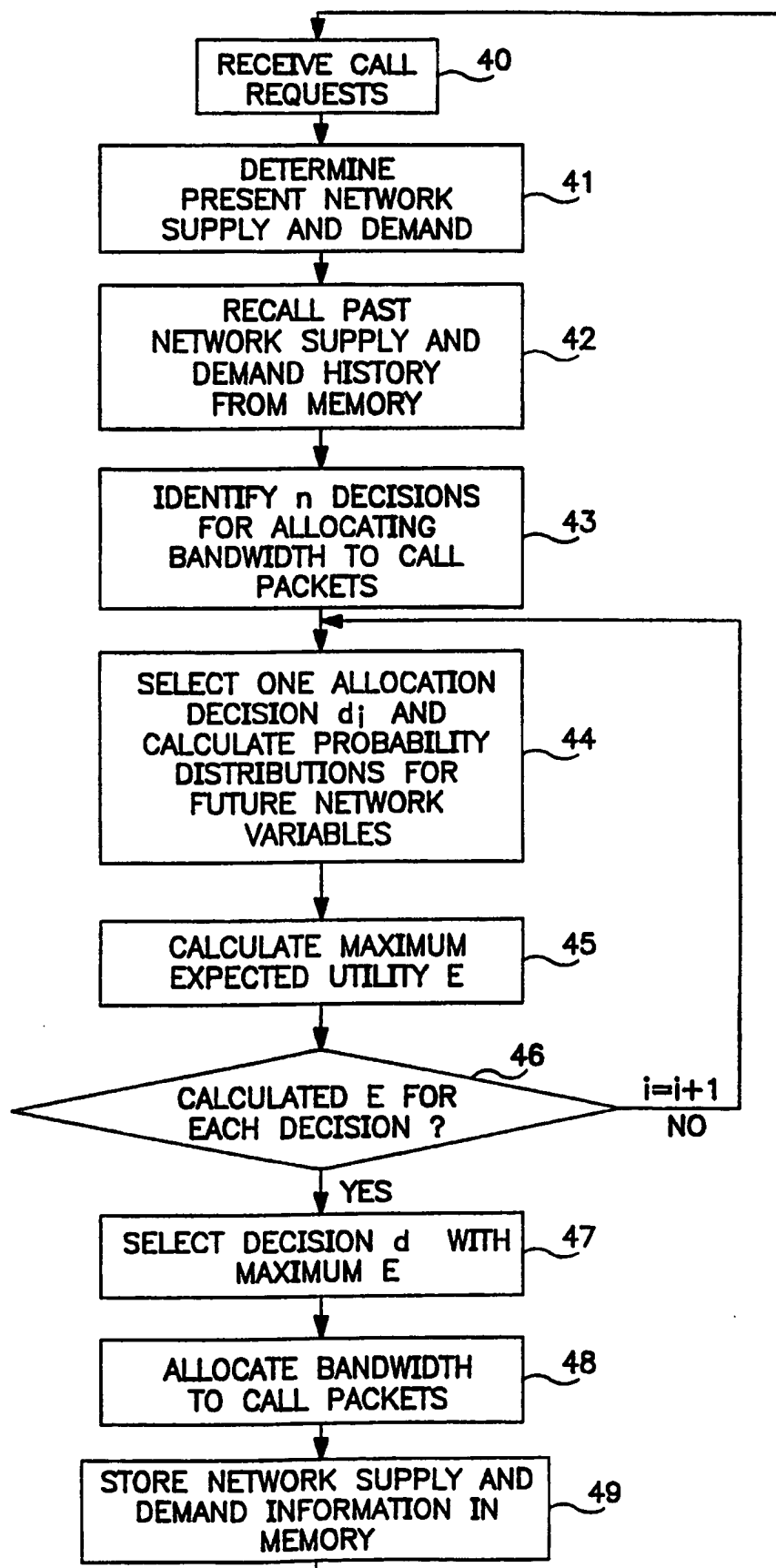


FIG. 4

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US98/21625

A. CLASSIFICATION OF SUBJECT MATTER

IPC(6) :H04B 10/105

US CL :455/12.1, 13.1, 450

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 455/12.1, 13.1, 450, 427, 428, 430, 451, 452, 453, 509

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

APS, Derwent

search terms: utility function, predicted, expected, future, allocate, assign

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y,P	US 5,402,478 A (HLUCHU ET AL.) 28 MARCH, 1998, COL 1-3, LINES 31-37 COL 16	1-2,17-18
Y	US 5,392,450 A (NOSSEN) 21 FEBRUARY 1995, LINES 19-29 COL 2, LINES 53-56 COL 11, AND ABSTRACT.	1-2,17-18
A,P	US 5,784,616 A (HORVITZ) 21 JULY, 1998 LINES 1-33 COL 25	1-31

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

* Special categories of cited documents:	"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
A document defining the general state of the art which is not considered to be of particular relevance	"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"B" earlier document published on or after the international filing date	"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"A" document member of the same patent family
"O" document referring to an oral disclosure, use, exhibition or other means	
"P" document published prior to the international filing date but later than the priority date claimed	

Date of the actual completion of the international search

17 DECEMBER 1998

Date of mailing of the international search report

25 FEB 1999

Name and mailing address of the ISA/US
Commissioner of Patents and Trademarks
Box PCT
Washington, D.C. 20231

Facsimile No. (703) 305-3230

Authorized officer

MAKOTO AOKI

Telephone No. (703) 308-9640